

*Int. J. Mol. Sci.* **2006**, *7*, 358–374

International Journal of  
**Molecular Sciences**

ISSN 1422-0067

© 2006 by MDPI

[www.mdpi.org/ijms/](http://www.mdpi.org/ijms/)

## Prediction of Environmental Properties for Chlorophenols with Posetic Quantitative Super-Structure/Property Relationships (QSSPR)

Teodora Ivanciuc <sup>1,\*</sup>, Ovidiu Ivanciuc <sup>2</sup> and Douglas J. Klein <sup>1,\*</sup>

<sup>1</sup> Texas A&M University at Galveston, Galveston, Texas 77553, USA; Email: [ivanciuc@netscape.net](mailto:ivanciuc@netscape.net); [kleind@tamug.edu](mailto:kleind@tamug.edu)

<sup>2</sup> Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas 77555-0857, USA; Email: [iejmd@yahoo.com](mailto:iejmd@yahoo.com)

\* Authors to whom correspondence should be addressed.

*Received: 31 May 2006 / Accepted: 14 August 2006 / Published: 28 September 2006*

---

**Abstract:** Due to their widespread use in bactericides, insecticides, herbicides, and fungicides, chlorophenols represent an important source of soil contaminants. The environmental fate of these chemicals depends on their physico-chemical properties. In the absence of experimental values for these physico-chemical properties, one can use predicted values computed with quantitative structure-property relationships (QSPR). As an alternative to correlations to molecular structure we have studied the super-structure of a reaction network, thereby developing three new QSSPR models (poset-average, cluster-expansion, and splinoid poset) that can be applied to chemical compounds which can be hierarchically ordered into a reaction network. In the present work we illustrate these poset QSSPR models for the correlation of the octanol/water partition coefficient ( $\log K_{ow}$ ) and the soil sorption coefficient ( $\log K_{OC}$ ) of chlorophenols. Excellent results are obtained for all QSSPR poset models to yield:  $\log K_{ow}$ ,  $r = 0.991$ ,  $s = 0.107$ , with the cluster-expansion QSSPR; and  $\log K_{OC}$ ,  $r = 0.938$ ,  $s = 0.259$ , with the spline QSSPR. Thus, the poset QSSPR models predict environmentally important properties of chlorophenols.

**Keywords:** chlorophenols (CPs), reaction network; Hasse diagram; partially ordered set; poset; poset-average; cluster-expansion; splinoid poset; quantitative structure-property relationships; QSPR; quantitative structure-activity relationships (QSAR); octanol/water partition coefficient ( $\log K_{ow}$ ); soil sorption coefficient ( $\log K_{OC}$ ).

---

## 1. Introduction

The widespread use of synthetic organic compounds in industry, agriculture, health care, and household is an important source of soil and water contamination. Other sources of contamination are accidental spills, hazardous waste disposal sites, storage tanks, or municipal landfills. To minimize the environmental impact of organic pollutants, the remediation of contaminated soil usually starts with the extraction of the pollutants into an aqueous phase, followed, if necessary, by other chemical or biological treatments. Knowledge of various physico-chemical properties of the organic pollutants is necessary for the design of these remediation processes [1,2]. Whenever the values for these physico-chemical properties are not experimentally available, various quantitative structure-property relationships (QSPR) have often been used to predict these properties.

The success of the soil remediation process for a particular organic compound depends on the distribution of that chemical in soil/water or soil/solvent systems. The partition of an organic pollutant between the water (hydrophobic) and organic (hydrophobic) phases is generally correlated with various properties, such as the water solubility  $S$  and the octanol/water partition coefficient  $K_{ow}$ . The environmental fate of organic compounds is also correlated with the soil adsorption partition coefficient  $K_{OC}$ . The modeling of these properties from structural parameters, with various QSPR models, has been investigated in many papers [3-12].

Phenol and its derivatives are common environmental contaminants [13-18], and most of them are known or suspected to be human carcinogens. Besides the fact that phenols give an unpleasant taste and odor to drinking water, they are powerful toxics for various biological processes. Due to their widespread use in industry, household, forest industries, and as disinfectants, chlorophenols represent an important source of soil contaminants [13,19-21]. The environmental fate (soil adsorption, water solubility, partition between soil and water, reaction rates) of these chemicals depends on their physico-chemical properties.

Many chemical compounds, derived from a common molecular skeleton, can be organized in formal reaction networks, such as substitution–reaction networks, having the mathematical structure of a partially ordered set (or poset) [22-31]. Thus poset substitution–reaction networks are a type of super-structure which then might be utilized in property modeling [26]. Using this reaction poset, we have recently developed highly predictive quantitative super-structural QSSPR models (poset-average, cluster-expansion, and splinoid poset), and applied these models for chlorobenzenes [26,32], methylbenzenes [26], methylcyclobutanes [31], and polychlorinated biphenyls [33].

Unlike the classical QSPR & QSAR (quantitative structure-property & -activity relationships), the reaction poset super-structure QSSPR and QSSAR models do not use conventional molecular descriptors to correlate physical, chemical, or biological properties. In the reaction poset approach, the molecular properties are predicted from a response framework generated by the super-structure of the substitution–reaction network. In the present work we apply these poset QSSPR models to predict the octanol/water partition coefficient ( $\log K_{ow}$ ) and the soil sorption coefficient ( $\log K_{OC}$ ) of chlorophenols. These fittings are here (favorably) validated via a leave-one-out procedure.

## 2. The Reaction Poset Diagram for Phenol Substitution

The poset super-structural QSSPR and QSSAR models make special use of the mathematical structure of a partially ordered set induced from a substitution–reaction network, when a molecular skeleton is subjected to successive steps of substitution. Starting from an unsubstituted compound, substituents are progressively introduced one after another, with earlier substituents fixed at their different possible positions.

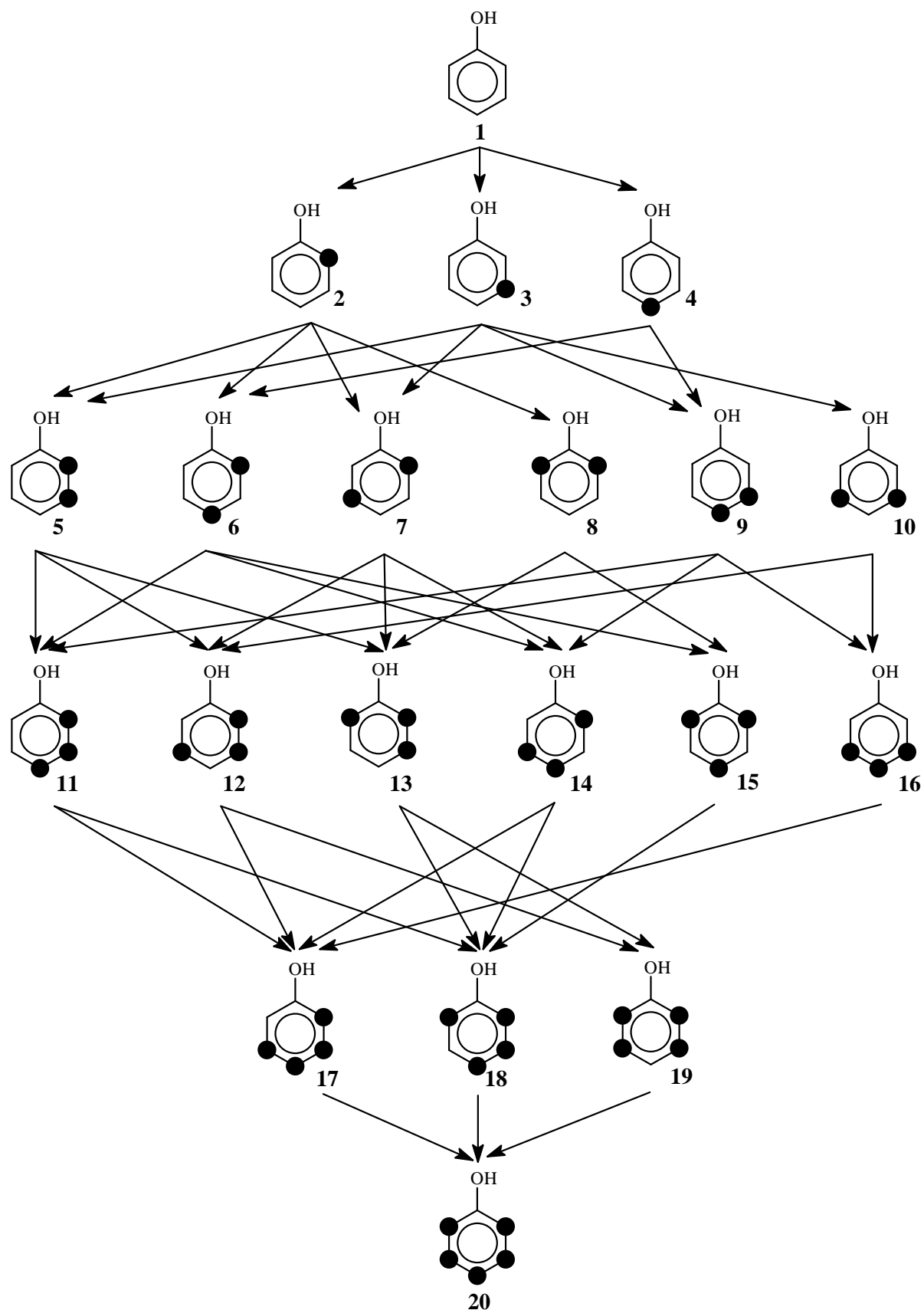
The special super-structure considered here is the substitution–reaction network that starts with phenol and continues with consecutive formal substitution reactions in which a H atom from the phenyl ring is replaced with a Cl atom (Figure 1). The poset reaction diagram starts with phenol at the top and ends with pentachlorophenol at the bottom, while all the remaining different patterns of substitution occur in between. The arrows indicate the hierarchic generation of the different patterns of more substituted compounds from the different patterns of less substituted ones.

As we present in detail in the following sections, the poset reaction diagram from Figure 1 is subjected to various mathematical treatments to generate poset super-structural QSSPR and QSSAR models to predict the octanol/water partition coefficient and the soil sorption coefficient of chlorophenols. The topology of the chlorophenol reaction poset is the basis for all these models, which is a notable departure from the classical QSAR and QSPR models that use various structural descriptors.

## 3. Experimental Data

As can be seen from Figure 1, chlorinated phenols constitute a series of 19 substituted compounds, which can be further classified as three monochlorophenols, six dichlorophenols, six trichlorophenols, three tetrachlorophenols, and one pentachlorophenol. The parent phenol is included in the poset as a 20th member. The United States Environmental Protection Agency (EPA) has classified chlorophenols as priority pollutants owing to their environmental toxicity. Due to their wide use in industry and the household (as bactericides, insecticides, herbicides, fungicides, and wood preservatives) [13-21], chlorophenols are easily released in the environment, either from direct use or accidental spillage. As a consequence, they cause severe environmental problems, being frequently detected in surface water, wastewater, soil, and sediments [17,34,35]. Exposure to chlorophenols can result in irritations of the respiratory tract and of the eyes. Higher doses can induce convulsions, shortness of breath, coma, or even death. The toxicity of chlorophenols is determined by the number and position of the Cl atoms, and by the concentration in a particular environmental compartment.

Due to their importance as environmental pollutants which can produce serious risks for human health, we have developed reaction poset super-structural QSSPR and QSSAR models for octanol/water partition coefficients ( $\log K_{ow}$ ) and soil sorption partition coefficients ( $\log K_{oc}$ ) of chlorophenols. All experimental data were collected from the literature:  $\log K_{ow}$  [36,37];  $\log K_{oc}$  [38].



**Figure 1.** The posetic phenol super-structural substitution–reaction network. The black enlarged dots indicate the sites on which an aromatic H atom of phenol has been replaced by a Cl atom.

## 4. Poset Methodology

### 4.1 Poset Applications in General

Partially ordered sets (or posets) have been advocated as of very general utility in chemistry [22,23], having numerous chemical applications [24]. Brüggemann and co-workers [39-44] have proposed their use as an attractive way of handling complex information within the environmental area. Poset models in ranking or prioritizing chemical pollutants have been proposed [45-48]. A book on the chemical and environmental science applications has appeared [49], and beyond this they are advocated [50] as of rather general utility in science, with there then also being numerous mathematical developments.

Formally a poset consists of a set  $P$  with a relation  $\succ$  which satisfies two conditions: first, for  $\alpha, \beta \in P$ ,  $\alpha \succ \beta \Rightarrow \beta \not\succ \alpha$ ; and second for  $\alpha, \beta, \gamma \in P$ ,  $\alpha \succ \beta$  and  $\beta \succ \gamma \Rightarrow \alpha \succ \gamma$ . In the particular case of chlorophenols (Figure 1) the set  $P$  consists of chemical compounds derived from phenol by substituting aromatic H atoms with Cl atoms, and the ordering  $\alpha \succ \beta$  means that  $\beta$  is obtainable from  $\alpha$  after some (non-zero) number of chlorinations. The relation which allows either  $\alpha \succ \beta$  or  $\alpha = \beta$  is denoted  $\alpha \succcurlyeq \beta$ , and the relation where  $\alpha \succ \beta$  without any intervening members of  $P$  is denoted  $\alpha \rightarrow \beta$ , in which case one says  $\alpha$  covers  $\beta$ . The Hasse diagram  $H(P)$  of  $P$  displays these covering relations, chosen to be oriented downward.

### 4.2 Reaction Poset Super-Structures

As presented in Figure 1, the chemical basis of our reaction-poset super-structural models is represented by the mathematical structure of a partially ordered set induced from a substitution–reaction network when a molecular skeleton is subjected to successive steps of substitution. The mathematical poset focused on here is represented just by the bare super-structural reaction network (or Hasse diagram), without explicit reference to the molecular structures shown at the different nodes of the network in Figure 1.

These reaction-network posets are of a special type. They always have a unique maximum and a unique minimum, and moreover each is self-dual, mapping into itself under the interchange of substituted and unsubstituted sites. Yet further these posets are ranked (according to the number of substituents), those members at the same rank being isomers. In general these posets are not mathematical lattices (defined as posets for which every pair of elements has a unique least upper bound and a unique greatest lower bound). In particular, our phenol substitution poset is not a lattice – e.g., because member **5** and **7** do not have a unique least upper bound (but rather two: **2** and **3**). But still they have an interesting structure, reminiscent of a “finite geometry” on a space of skeletal substitution positions, with the geometric structure mediated by the skeletal group, here  $C_6H_5OH$  for our phenol example.

### 4.3 Poset QSSPR and QSSAR Modelling

The reaction poset super-structure QSPR models considered here are based on the substitution–reaction network that starts with phenol and continues with consecutive formal substitution reactions in which a H atom from the phenyl ring is replaced with a Cl atom. After five steps of successive

substitution, all reaction branches converge to pentachlorophenol, which concludes the reaction network (Figure 1). Each vertex in the Hasse diagram may be identified to the property value for the corresponding compound.

The topology of the chlorophenol reaction poset is the basis for all models investigated in the present paper, namely poset-average, cluster-expansion, and splinoid poset. Otherwise information about the molecular structures is foregone – though it may be seen that the poset has embedded in it much information about molecular structure, and especially about interrelations between molecular structures. Following our previous procedure tested for a number of chemical classes (chlorobenzenes [26,32], methylbenzenes [26], methylcyclobutanes [31], and polychlorinated biphenyls [33]), we evaluate the models by comparing their leave-one-out (LOO) cross-validation statistics giving them the correlation coefficient  $r$  and standard deviation  $s$ . We next briefly describe the three reaction poset super-structural QSSPR models to be utilized here.

#### 4.4 Average-Poset Model

Starting from the Hasse diagram (Figure 1) our poset-average method [26] computes a predicted value  $X(\beta)_{\text{pred}}$  for a property  $X$  of a compound  $\beta$  as the average of two averages, namely the average of experimental values  $X(\alpha)_{\text{exp}}$  for all compounds  $\alpha$  from the previous level that connect by incoming arrows to  $\beta$ , and the average of experimental values  $X(\gamma)_{\text{exp}}$  for all compounds  $\gamma$  from the next level that receive outgoing arrows from  $\beta$ . To apply this the experimental property values must be available for all diagram positions adjacent to  $\beta$ . For example, in Figure 2 we present the reaction poset diagram for chlorophenols, in which each vertex (compound) has attached the experimental value for  $\log K_{\text{ow}}$  [36,37]. The poset-average  $\log K_{\text{ow}}$  predicted value for 4-chlorophenol (4-CIP) is computed with the formula:

$$\begin{aligned}\log K_{\text{ow}}(4\text{-CIP}) &= \frac{1}{2} \left\{ \log K_{\text{ow}}(\text{P}) + \frac{1}{2} [\log K_{\text{ow}}(2,4\text{-Cl}_2\text{P}) + \log K_{\text{ow}}(3,4\text{-Cl}_2\text{P})] \right\} \\ &= \frac{1}{2} \left\{ 1.46 + \frac{1}{2} [3.22 + 3.17] \right\} = 2.33\end{aligned}$$

As one can see from this example, the properties computed with the poset-average method are parameter-free predictions, and the statistical indices are obtained via LOO statistics.

#### 4.5 Splinoid Poset Model

The chloro-substitution network of phenol is represented here as a Hasse diagram  $H(P)$  (Figure 1) which mathematically represents a finite poset  $P$ . An oriented edge in the Hasse diagram here represents the transition  $\alpha \rightarrow \beta$  from a chemical compound  $\alpha$  with  $n$  chlorine atoms to one  $\beta$  with  $n+1$  chlorine atoms, and we attach a real variable  $x_{\alpha \rightarrow \beta}$  ranging from 0 to 1, that represents the transformation of  $\alpha$  into  $\beta$ . When formulating the splinoid QSSPR model for a property  $X$ , one considers cubic spline polynomials (in  $x_{\alpha \rightarrow \beta}$ ) on the oriented edges  $\alpha \rightarrow \beta$  of the Hasse diagram  $H(P)$ . Further each vertex  $\alpha$  of  $H(P)$  or  $P$  is identified by a value  $a_\alpha$  and a slope  $b_\alpha$  for the spline polynomials incident at  $\alpha$ . The splinoid poset QSSPR model is generated based on known values of the property  $X$  for a subset  $K \subseteq P$  of the chemical compounds. Briefly, the splinoid fit consists of the following steps: first, the cubic splines match values  $a_\alpha$  at the nodes  $\alpha \in K$  to the known property values; second, the

incoming and outgoing slopes through each node match to the corresponding  $b_\alpha$  value; and third, a relevant total “curvature” of the overall spline fit is minimized (subject to the constraints of the first two conditions). With the splinoid QSSPR determined for the vertices from  $K$ , one can predict the property values for the remaining chemical compounds that do not have an experimental value for the property  $X$  these being the compounds that form the “unknown” set  $U$  of vertices  $\alpha \notin K$ .

A mathematical derivation [27] leads to a closed formula predicting the values of  $X$  for the set  $U$  of chemical compounds. Let  $\mathbf{A}$  denote the adjacency matrix of the Hasse diagram  $H(P)$ , and let  $\mathbf{S}$  denote the oriented adjacency matrix of  $H(P)$ , where:

$$S_{\alpha\beta} = \begin{cases} 1 & , \text{if } \beta \rightarrow \alpha \\ -1 & , \text{if } \alpha \rightarrow \beta \\ 0 & , \text{otherwise} \end{cases}$$

The in-degree on vertex  $\alpha \in P$  is denoted by  $d_{\rightarrow\alpha}$  and the out-degree on vertex  $\alpha \in P$  is denoted by  $d_{\alpha\rightarrow}$ . Then, we introduce two diagonal matrices:

$$\begin{aligned} \mathbf{D} &= \text{diag}[d_{\alpha\rightarrow} - d_{\rightarrow\alpha}] \\ \mathbf{\Lambda} &= \text{diag}[d_{\alpha\rightarrow} + d_{\rightarrow\alpha}] \end{aligned}$$

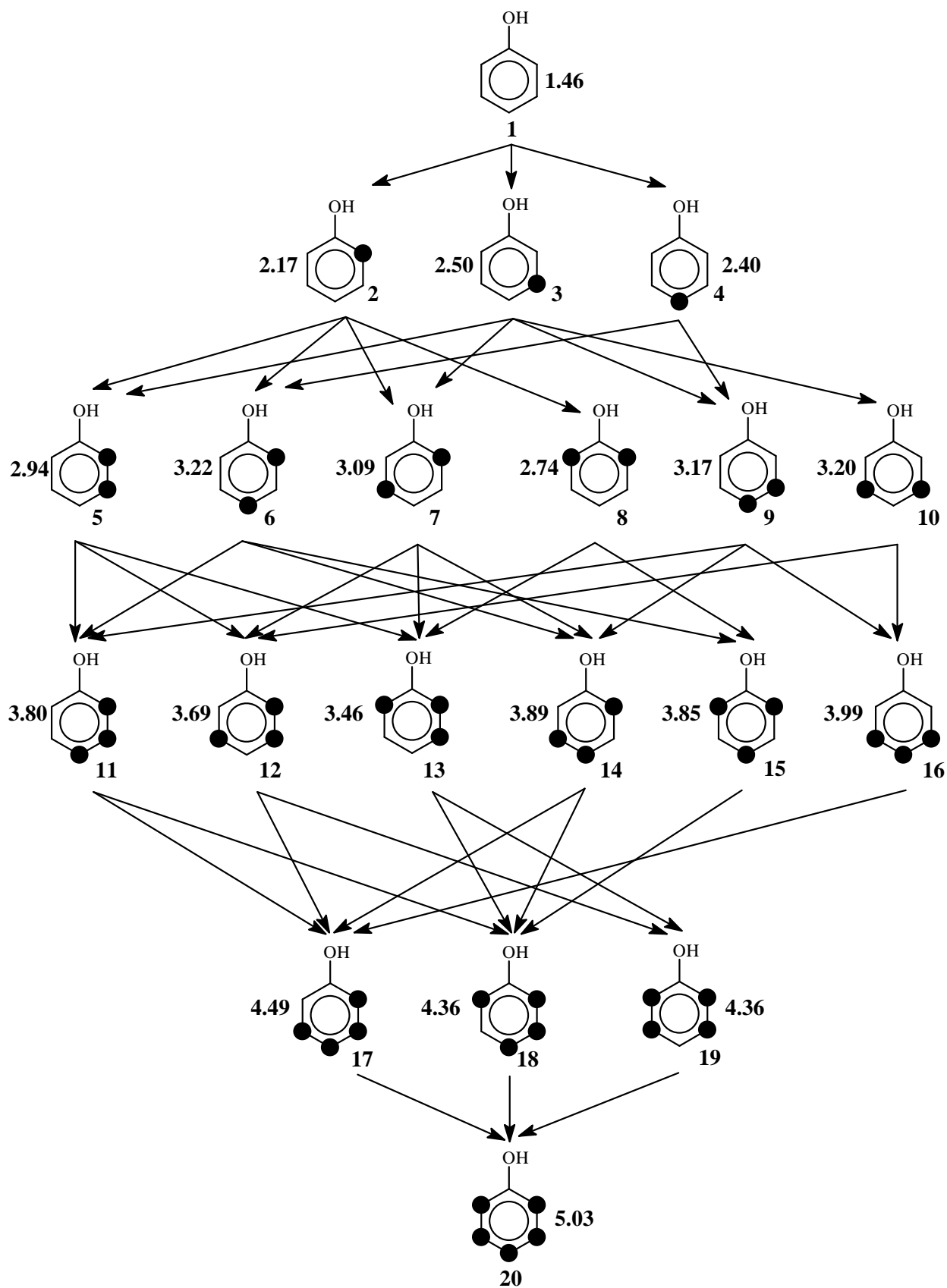
Further define the matrices  $\mathbf{U}$  (the  $|U| \times |P|$  submatrix of the unity matrix  $\mathbf{I}$ , with rows indexed by the elements of  $U$ ), and  $\mathbf{K}$  (the  $|K| \times |P|$  submatrix of the unity matrix  $\mathbf{I}$ , with rows indexed by the elements of  $K$ ), and the derived matrix:

$$\mathbf{M} = 2(\mathbf{\Lambda} - \mathbf{A}) - 3(\mathbf{D} - \mathbf{S})(\mathbf{A} + 2\mathbf{\Lambda})^{-1}(\mathbf{D} + \mathbf{S})$$

The (column) vector of known property values is denoted by  $\vec{k}$ . Then, the vector  $\vec{u}$  that contains the predictions for the unknown property values  $a_\alpha$  is computed from:

$$\vec{u} = -(\mathbf{UMU}^t)^{-1}(\mathbf{UMK}^t)\vec{k}$$

For a few different reaction networks we have studied the matrix  $\mathbf{UMU}^t$  which appears in practice to be invertible regardless of how sparse the “known” data is in the network up to the point that very few ( $\leq 2$ ) known data are available. The coefficients appearing in the spline polynomials do not explicitly appear in our splinoid formula for  $\vec{u}$ , but they are complicit in the derivation of this formula for  $\vec{u}$ . The present formula gives  $\vec{u}$  in terms of the poset structure, and thence completes the splinoid QSSPR algorithm, which turns out to give a robust model in accommodating a diversity of missing values for several compounds (which may possibly even be adjacent). This is a significant advantage of the splinoid model, which uses the topology of the Hasse diagram to generate a response network for the investigated property. To achieve comparison with the results from the other poset QSSPR models, we have used the splinoid model in the leave-one-out cross-validation procedure.



**Figure 2.** The reaction poset diagram of chlorophenols with the experimental values of the octanol/water partition coefficients ( $\log K_{ow}$ ) [36,37].



#### 4.6 Cluster-Expansion Model

Formal cluster-expansion in general re-expresses a scalar function (or property) for the different members of a poset in terms of related functions focusing more strongly on earlier members of the poset. Much of the formal theory is described by Rota [51] for general posets, and its chemical application in the case that the partial ordering is the subgraph partial ordering is described in [28-31]. Generally, for a scalar property  $X$  defined on the members of a poset  $P$  (with partial ordering  $\succ$ ) one may expand  $X$  for  $\alpha \in P$ , as

$$X(\alpha) = \sum_{\beta \preceq \alpha} f(\beta, \alpha) X_f(\beta)$$

where the sum goes over all  $\beta \preceq \alpha$ ,  $f(\beta, \alpha)$  is a *cluster* function that maps pairs of members of  $P$  onto real numbers with  $f(\beta, \alpha) = 0$  whenever  $\beta \not\preceq \alpha$ , and is such that  $f(\alpha, \alpha) \neq 0$ . Further,  $X_f(\beta)$  is an *f transform* property depending on  $X$  and the cluster function  $f$ . Conveniently, this cluster-expansion may be truncated to a limited sequence of non-zero cluster approximants, and so applied whenever the earlier terms offer a good approximation of the property  $X$ .

For our reaction-network posets, we choose [31-33] that  $f(\beta, \alpha)$  be the number of ways in which substitution pattern  $\alpha$  occurs as a subset of substitution pattern  $\beta$ . For the poset diagram of chlorophenols, we have truncated the cluster-expansion model to  $X_f$  contributions from the chlorine atoms situated through the second and third rows of the poset (Figure 1). The number of parameters (*i.e.*, the  $X_f(\beta)$ ) from the third row is reduced from 5 to 3 through the approximation of making them depend solely on the relative positions of the two chlorine atoms (as *ortho*, *meta*, and *para*):

$$\begin{aligned} X_f(2,3\text{-Cl}_2\text{P}) &= X_f(3,4\text{-Cl}_2\text{P}) \equiv d \\ X_f(2,4\text{-Cl}_2\text{P}) &= X_f(3,5\text{-Cl}_2\text{P}) \equiv e \\ X_f(2,5\text{-Cl}_2\text{P}) &\equiv f \end{aligned}$$

where P indicates phenol. The parameters associated to the second row of the poset are abbreviated to

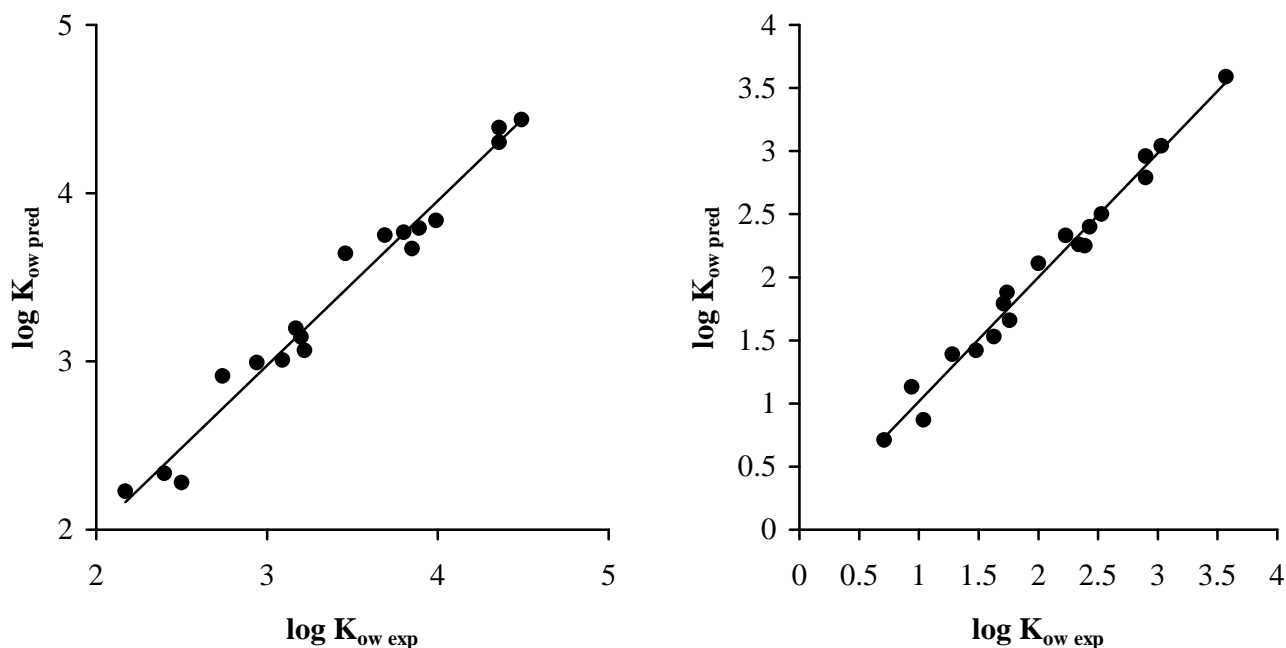
$$X_f(2\text{-ClP}) \equiv a, X_f(3\text{-ClP}) \equiv b, X_f(4\text{-ClP}) \equiv c.$$

This truncated cluster-expansion model proves to be able to model the properties of chlorophenols. In each series of QSSPR models, phenol was considered as a reference structure, namely, the property values are shifted so that  $X(\text{phenol}) = 0$ . The set of  $X_f(\beta)$  parameters ( $a, b, c, d, e, f$ ) can be computed by a least-squares procedure based on a subset of molecules, or by “inversion” from small systems - and here we use the former choice. All models were tested in a leave-one-out cross-validation procedure, in order to obtain results comparable with those from the other poset QSSPR models.

## 5. Results and Discussions

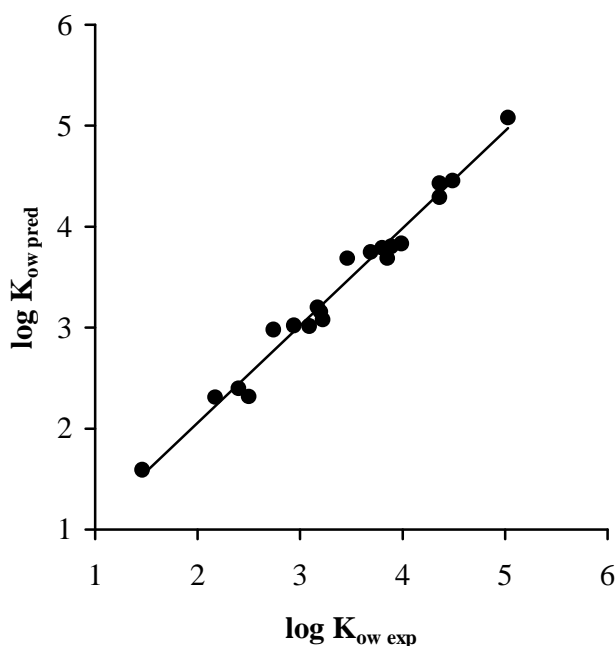
The first group of poset QSSPR models is developed for the octanol/water partition coefficient  $K_{ow}$  of chlorophenols. All 20 values, including that for phenol, were collected from the literature [36,37]. The predictions obtained with the reaction poset super-structure QSSPR models are of very good quality: poset-average,  $r = 0.987$ ,  $s = 0.115$ ; cluster-expansion,  $r = 0.991$ ,  $s = 0.107$ ; splinoid poset,  $r =$

0.990,  $s = 0.122$ . As can be seen from the plots of experimental vs. predicted  $K_{ow}$  values (Figure 3), there are no significant outliers or deviations from linearity.



average method:  $r = 0.987$ ;  $s = 0.115$

cluster-expansion method:  $r = 0.991$ ;  $s = 0.107$



spline method:  $r = 0.990$ ;  $s = 0.122$

**Figure 3.** Plot of experimental vs. predicted octanol/water partition coefficient for chlorophenols with the poset-average, cluster-expansion, and splinoid poset QSSPR models.

The second application for  $\log K_{OC}$  considers the situation when not all 20 experimental values of the chlorophenols are known. We found in the literature only 12 values for the soil sorption coefficient  $K_{OC}$  for chlorophenols and phenol [38]. Due to the absence of a significant number of experimental

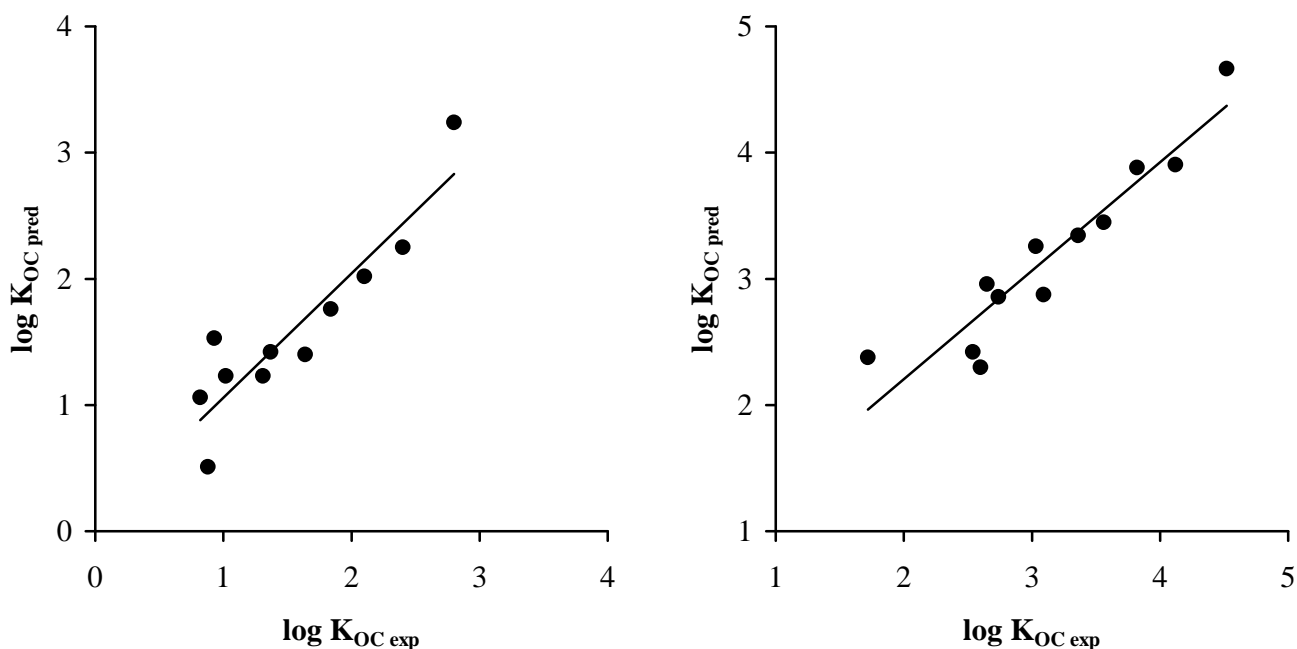
values, the poset-average method cannot be used. On the other hand, we obtained good statistics for the cluster-expansion ( $r = 0.912$ ,  $s = 0.287$ ) and splinoid poset ( $r = 0.938$ ,  $s = 0.259$ ). The predictive values by these two different methods are identified in Table 1. The splinoid scheme reproduces exactly the 12 known experimental values, which then in the Table 1 are entered in bold-face. Comparison of predictions for the 12 known ones when one-by-one they are left out are shown in Figure 4.

**Table 1.** Experimental and predicted values for cluster-expansion and splinoid QSSPR models for soil sorption coefficient,  $\log K_{OC}$ . The experimental values are presented in bold.

No.	Compound	Splinoid model	Cluster-expansion
1	P	<b>1.72</b>	1.72
2	2-ClP	<b>2.60</b>	2.23
3	3-ClP	<b>2.54</b>	2.78
4	4-ClP	2.42	2.81
5	2,3-Cl <sub>2</sub> P	<b>2.65</b>	3.25
6	2,4-Cl <sub>2</sub> P	<b>2.74</b>	2.95
7	2,5-Cl <sub>2</sub> P	2.87	2.68
8	2,6-Cl <sub>2</sub> P	2.78	2.51
9	3,4-Cl <sub>2</sub> P	<b>3.09</b>	3.14
10	3,5-Cl <sub>2</sub> P	2.92	2.71
11	2,3,4-Cl <sub>3</sub> P	3.32	3.43
12	2,3,5-Cl <sub>3</sub> P	3.35	3.09
13	2,3,6-Cl <sub>3</sub> P	3.24	2.99
14	2,4,5-Cl <sub>3</sub> P	<b>3.36</b>	3.12
15	2,4,6-Cl <sub>3</sub> P	<b>3.03</b>	2.95
16	3,4,5-Cl <sub>3</sub> P	<b>3.56</b>	3.48
17	2,3,4,5-Cl <sub>4</sub> P	<b>4.12</b>	3.97
18	2,3,4,6-Cl <sub>4</sub> P	<b>3.82</b>	3.74
19	2,3,5,6-Cl <sub>4</sub> P	3.92	3.58
20	Cl <sub>5</sub> P	<b>4.52</b>	4.96

Overall the correlation coefficients are very good for such complex property correlations, whence a subsequent natural question concerns the relation to molecular structure and a comparison to more conventional QSPR fittings. There are many hundreds of possible choices of molecular-structure descriptors, so that a definitive comparison to QSPR is elusive, even for the limited case of chlorophenols, though the more fundamental question concerns a more general range. But obviously QSPR schemes focus on molecular structure as the fundamental object of study, whereas our posetic approach focuses on the super-structural reaction network as the fundamental object of study (so that we have used the abbreviation QSSPR). Questions of what QSSPR tells us about molecular structure, though rather incompletely answered, might be compared to the incompletely answered converse question of what ordinary QSPR approaches tell one about the reaction network. Our splinoid QSSPR approach clearly tends to assign similar values for structures which are similar in the sense that they have a large common graphical substructure (since then the two molecular structures are close together

in the reaction poset), while the splinoid fit interpolates as smoothly as possible between the nearby known values. Likewise with two molecular structures sharing a large common substructure and so being nearby in the posetic diagram, the cluster expansion we make gives a similar set of predecessors for two such nearby structures, and thence similar numerical values for the fitted property. Both QSPR & QSSPR schemes, then tend to assign similar property values to "similar" structures. We believe that there is an even tighter formal relationship between our reaction-network cluster expansion and common (QSPR-based) substructural cluster expansions – as is seen in the examples where we have indicated a molecular substructural interpretation of our retained reaction-network-cluster terms. We believe there is a general correspondence between the two types of cluster expansions, though in the structural & super-structural circumstances the terms are ordered differently, and thence different later terms are generally omitted in the two schemes. This surely warrants more formal study, only the beginning of which is described in [31], and is not pursued here.



cluster-expansion method:  $r = 0.912$ ;  $s = 0.287$

spline method:  $r = 0.938$ ;  $s = 0.259$

**Figure 4.** Plot of experimental vs. predicted soil sorption coefficient for chlorophenols with the cluster-expansion and splinoid poset QSSPR models.

Overall it seems that one might frequently anticipate similar fittings from QSPR & QSSPR schemes – so long as the QSPR is limited to structures occurring within the reaction-network superstructure. As an example comparison, we consider QSPR fittings to two structural indexes:  $\#_{\text{Cl}}(\xi)$  the number of chlorine atoms in the chlorophenol  $\xi$ ; and  $\chi(\xi)$  the Randic connectivity index for the H-deleted graph (not distinguishing C, O, or Cl atoms). That is, one considers a fitting of a molecular property  $X$  to

$$X(\xi) = A + B \cdot \#_{\text{Cl}}(\xi) + C \cdot \chi(\xi)$$

We make two least-squares fittings, for our two presently studied properties. The results for  $\log K_{ow}$  are:

$$A = 39.484, B = 5.3414, C = -11.207 \quad \text{with} \quad r = 0.993, s = 0.109, r_{LOO} = 0.991$$

the statistics here being excellent. For  $\log K_{OC}$  the results are:

$$A = -3.325, B = -0.1135, C = 1.530 \quad \text{with} \quad r = 0.966, s = 0.221, r_{LOO} = 0.943$$

which also are very good statistics. As expected from the excellence of our earlier cluster-expansion fit, and its close relationship to typical invariants for QSPR fittings, the results for either type of approach are very good, and very similar as to error statistics. Though the results are comparable, what we have done is to show that an alternative novel sort of (QSSPR) approach is also available, and that for the example here along with a few elsewhere, rather high quality fits are achievable.

### 3. Conclusions

Chlorophenols are widely used as bactericides, insecticides, herbicides, fungicides and wood preservatives [13-21], which makes them frequent environmental pollutants, either from direct use or accidental spillage. Exposure to chlorophenols can result in irritations of the respiratory tract and of the eyes. Commonly detected in surface water, wastewater, soil, and sediments [17,34,35], chlorophenols were classified by the EPA as priority pollutants. The investigation of their sorption behavior is fundamental to simulate and eventually predict their environmental fate. Because the octanol/water partition coefficient  $K_{ow}$  and the soil sorption partition coefficient  $K_{OC}$  are useful to estimate the mobility of an organic compound in soil, both are important to understand the distribution of chemical compounds in soil, sediments, and water. Because the laboratory methods for the determination of  $K_{ow}$  and  $K_{OC}$  are time consuming, the reaction poset super-structure QSSPR models demonstrated here can be applied to obtain reliable predictions for these properties.

To predict the octanol/water partition coefficient  $\log K_{ow}$  and the soil sorption coefficient  $\log K_{OC}$  of chlorophenols we have compared the predictive power of three reaction poset super-structural QSSPR models developed in our group [22-33], namely poset-average, cluster-expansion, and splinoid poset. The poset super-structural QSSPR models make special use of the mathematical structure of a partially ordered set induced in a substitution–reaction network when a molecular skeleton (such as benzene, naphthalene, or biphenyl) is subjected to successive steps of substitution. Starting from an unsubstituted compound, substituents are progressively introduced one after another, with earlier substituents fixed at their different possible positions. The special super-structure considered here is the substitution–reaction network that starts with phenol and continues with consecutive formal substitution reactions in which a H atom from the phenyl ring is replaced with a Cl atom. The poset reaction diagram starts with phenol at the top and ends with pentachlorophenol at the bottom, while all the different patterns of substitution occur in between. The poset-average is a local non-parametric method, the cluster-expansion is a parametric method, and the splinoid poset method is a global interpolation method.

Based on the poset reaction diagram, all three of these QSSPR models reflect in distinct ways the topology of the network that describes the interconversion of chemical species. All three poset QSSPR

methods give very good predictions for the properties investigated here. For  $\log K_{ow}$ , the cluster-expansion gives slightly better leave-one-out predictions & validations ( $r = 0.991$ ,  $s = 0.107$ ), while for  $\log K_{oc}$  the best LOO predictions & validations are obtained with the splinoid poset method ( $r = 0.938$ ,  $s = 0.259$ ). Thus, we have extended the application of the poset QSSPR models to the prediction of environmentally important properties of chlorophenols. Evidently especially the splinoid and cluster-expansion models are applicable to circumstances where there is missing data, as in the case of the soil sorption coefficient. There seems promise for further similar uses of such posetic reaction-networks for QSSPR and QSSAR modeling. But in addition, it seems to us that it would be of value to further extend our approach with the simultaneous use of two or more reactions, so as to treat in one setting a larger range of structures – this then yielding a "multi-poset". Further, we think that it could be interesting if there were revealed a formal relation between QSSAR (or QSSPR) on one hand and QSAR (or QSPR) on the other. In particular, it would be of interest if features of the present QSSPR (or QSSAR) were identified to engender greater distinction in fittings. Certainly much work remains, both in the general context of partial orderings, and for our currently studied special case of substitution-reaction-network posets.

### Acknowledgements

The authors acknowledge the support (through grant BD-0894) of the Welch Foundation of Houston, Texas.

### References and Notes

1. Kaise, K. L. E. Organic Contaminants in the Environment: Research Progress and Needs. *Environ. Intern.* **1984**, *10*, 241-250.
2. Kieth, L. H.; Telliard, W. A. Priority Pollutants, Part I: A Perspective View. *Envir. Sci. Technol.* **1979**, *13*, 416-423.
3. Guo, R.; Liang, X.; Chen, J.; Wu, W.; Zhang, Q.; Martens, D.; Kettrup, A. Prediction of Soil Organic Carbon Partition Coefficients by Soil Column Liquid Chromatography. *J. Chromatogr. A* **2004**, *1035*, 31–36.
4. Gramatica, P.; Corradi, M.; Consonni, V. Modelling and Prediction of Soil Sorption Coefficients of Non-Ionic Organic Pesticides by Molecular Descriptors. *Chemosphere* **2000**, *41*, 763-777.
5. Chu, W.; Chan, K. H. The Prediction of Partitioning Coefficients for Chemicals Causing Environmental Concern. *Sci. Tot. Environ.* **2000**, *248*, 1-10.
6. Chiou, C. T.; Schmedding, D. W.; Manes, M. Improved Prediction of Octanol-Water Partition Coefficients from Liquid-Solute Water Solubilities and Molar Volumes. *Environ. Sci. Technol.* **2005**, *39*, 8840-8846.
7. Sabljíć, A.; Güsten, H.; Verhaar, H.; Hermens, J. QSAR Modelling of Soil Sorption. Improvements and Systematics of  $\log K_{oc}$ , vs.  $\log K_{ow}$  Correlations. *Chemosphere* **1995**, *31*, 4489-4514.
8. Briggs, G. G. Theoretical and Experimental Relationships Between Soil Adsorption, Octanol-Water Partition Coefficients, Water Solubilities, Bioconcentration Factors, and the Parachor. *J. Agric. Food Chem.* **1981**, *29*, 1050-1059.

9. Sabljíć, A. On the Prediction of Soil Sorption Coefficients of Organic Pollutants from Molecular Structure: Application of Molecular Topology Model. *Environ. Sci. Technol.* **1987**, *21*, 358-366.
10. Baker, J. R.; Mihelcic, J. R.; Sabljíć, A. Reliable QSAR for Estimating  $K_{oc}$  for Persistent Organic Pollutants: Correlation with Molecular Connectivity Indices. *Chemosphere* **2001**, *45*, 213-221.
11. Yang, K.; Zhu, L.; Lou, B.; Chen, B. Correlations of Nonlinear Sorption of Organic Solutes with Soil/Sediment Physicochemical Properties. *Chemosphere* **2005**, *61*, 116-128.
12. Chiou, C. T.; Schmedding, D. W.; Manes, M. Partitioning of Organic Compounds in Octanol-Water Systems. *Environ. Sci. Technol.* **1982**, *16*, 4-10.
13. <http://www.atsdr.cdc.gov/toxprofiles/tp107.html>; Toxicological Profile for Chlorophenols.
14. Nichkova, M.; Marco, M. P. Biomonitoring Human Exposure to Organohalogenated Substances by Measuring Urinary Chlorophenols Using a High-Throughput Screening (HTS) Immunochemical Method. *Environ. Sci. Technol.* **2006**, *40*, 2469-2477.
15. Zhao, F.; Mayura, K.; Hutchinson, R.W.; Lewis, R. P.; Burghardt, R. C.; Phillips, T. D. Developmental Toxicity and Structure-Activity Relationships of Chlorophenols Using Human Embryonic Palatal Mesenchymal Cells. *Toxicol. Lett.* **1995**, *78*, 35-42.
16. Chen, J.; Zhang, F.; Yu, F.; Zhang, J. Cytotoxic Effects of Environmentally Relevant Chlorophenols on L929 Cells and Their Mechanisms. *Cell Biol. Toxicol.* **2004**, *20*, 183-196.
17. Wang, Y. -J.; Lee, C. -C.; Chang, W. -C.; Liou, H. -B.; Ho, Y. -S. Oxidative Stress and Liver Toxicity in Rats and Human Hepatoma Cell Line Induced by Pentachlorophenol and its Major Metabolite Tetrachlorohydroquinone. *Toxicol. Lett.* **2001**, *122*, 157-169.
18. Kukkonen, J. V. K. Lethal Body Residue of Chlorophenols and Mixtures of Chlorophenols in Benthic Organisms, *Arch. Environ. Contam. Toxicol.* **2002**, *43*, 214-220.
19. Buikema, A. L. Jr.; McGinniss, M. J.; Cairns, J. Jr. Phenolics in Aquatic Ecosystems: A Selected Review of Recent Literature. *Marine Environ. Res.* **1979**, *2*, 87-181.
20. Ahlborg, U. G.; Thunberg, T. M. Chlorinated Phenols: Occurrence, Toxicity, Metabolism, and Environmental Impact. *CRC Crit. Rev. Toxicol.* **1980**, *7*, 1-35.
21. Czaplicka, M. Sources and Transformations of Chlorophenols in the Natural Environment. *Sci. Tot. Environ.* **2004**, *322*, 21-39.
22. Klein, D. J. Similarity and Dissimilarity in Posets. *J. Math. Chem.* **1995**, *18*, 321-348.
23. Klein, D. J.; Babić, D. Partial Orderings in Chemistry. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 656-671.
24. Klein, D. J. Prolegomenon on Partial Orderings in Chemistry. *Commun. Math. Comput. Chem. (MATCH)* **2000**, *42*, 7-21.
25. Klein, D. J.; Bytautas, L. Directed Reaction Graphs as Posets. *Commun. Math. Comp. Chem. (MATCH)* **2000**, *42*, 261-289.
26. Ivanciuc, T.; Klein, D. J. Parameter-Free Structure-Property Correlation via Progressive Reaction Posets for Substituted Benzenes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 610-617.
27. Došlić, T.; Klein, D. J. Splinoid Interpolation on Finite Posets. *J. Comput. Appl. Math.* **2005**, *177*, 175-185.
28. Klein, D. J. Chemical Graph-Theoretic Cluster Expansions. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1986**, *20*, 153-171.

29. Schmalz, T. G.; Živković, T.; Klein, D. J. Cluster Expansion of the Hückel Molecular Energy for Acyclics: Applications to pi Resonance Theory. *Math Chem. Comp.* **1987**, *54*, 173-190.
30. Klein, D. J.; Schmalz, T. G.; Bytautas, L. Chemical Sub-Structural Cluster Expansions for Molecular Properties. *SAR QSAR Environ. Res.* **1999**, *10*, 131-156.
31. Ivanciuc, T.; Klein, D. J.; Ivanciuc, O. Posetic Cluster Expansion for Substitution-Reaction Diagrams and its Application to Cyclobutane. *J. Math. Chem.* **2006**, web release.
32. Ivanciuc, T.; Ivanciuc, O.; Klein, D. J. Posetic Quantitative Superstructure/Activity Relationships (QSSARs) for Chlorobenzenes. *J. Chem. Inf. Model.* **2005**, *45*, 870-879.
33. Ivanciuc, T.; Ivanciuc, O.; Klein, D.J. Modeling the Bioconcentration Factors and Bioaccumulation Factors of Polychlorinated Biphenyls with Posetic Quantitative Super-Structure/Activity Relationships (QSSAR). *Molecular Diversity* **2006**, *10*, 133-145.
34. Arcand, Y.; Hawari, J.; Guiot, S. R. Solubility of Pentachlorophenol in Aqueous Solutions: The pH Effect. *Wat. Res.* **1995**, *29*, 131-136.
35. Diez, M. C.; Mora, M. L.; Videla, S. Adsorption of Phenolic Compounds and Color from Bleached Kraft Mill Effluent Using Allophanic Compounds. *Water Res.* **1999**, *33*, 125-130.
36. Xie, T. M.; Hulthe, B.; Folestad, S. Determination of Partition Coefficients of Chlorinated Phenols, Guaiacols and Catechols by Shake-Flask GC and HPLC. *Chemosphere* **1984**, *13*, 445-459.
37. Shiu, W.-Y.; Ma, K.-C.; Varhaníčková, D.; Mackay, D. Chlorophenols and Alkylphenols: A Review and Correlation of Environmentally Relevant Properties and Fate in an Evaluative Environment. *Chemosphere* **1995**, *29*, 1155-1224.
38. Saçan, M. T.; Balcioglu, I. A. Prediction of the Soil Sorption Coefficient of Organic Pollutants by the Characteristic Root Index Model. *Chemosphere* **1996**, *32*, 1993-2001.
39. Brüggemann, R.; Schwaiger, J.; Negele, R. D. Applying Hasse Diagram Technique for the Evaluation of Toxicological Fish Tests. *Chemosphere* **1995**, *30*, 1767-1780.
40. Brüggemann, R.; Bartel, H. G. A Theoretical Concept to Rank Environmentally Significant Chemicals. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 211-217.
41. Pudenz, S.; Brüggemann, R.; Luther, B.; Kaune, A.; Kreimes K. An Algebraic/Graphical Tool to Compare Ecosystems with Respect to Their Pollution V: Cluster Analysis and Hasse Diagrams. *Chemosphere* **2000**, *40*, 1373-1382.
42. Brüggemann, R.; Münzer, B.; Halfon, E. An Algebraic/Graphical Tool to Compare Ecosystems with Respect to Their Pollution - The German River "Elbe" as an Example - I: Hasse-Diagrams. *Chemosphere* **1994**, *28*, 863-872.
43. Brüggemann, R.; Pundez, S.; Carlsen, L.; Sørensen, P. B.; Thomsen, M.; Mishra, R. K. The Use of Hasse Diagrams as a Potential Approach for Inverse QSAR. *SAR QSAR Environ. Res.* **2001**, *11*, 473-487.
44. Carlsen, L.; Sørensen, P. B.; Thomsen, M.; Brüggemann R. QSAR's Based on Partial Order Ranking. *SAR QSAR Environ. Res.* **2002**, *13*, 153-165.
45. Lerche, D.; Sørensen, P. B.; Larsen H. L.; Carlsen, L.; Nielsen, O. J. Comparison of the Combined Monitoring-Based and Modelling-Based Priority Setting Scheme with Partial Order



- Theory and Random Linear Extensions for Ranking of Chemical Substances, *Chemosphere* **2002**, *49*, 637-649.
46. Lerche, D.; Sørensen, P. B. Evaluation of the Ranking Probabilities for Partial Orders Based on Random Linear Extension. *Chemosphere* **2003**, *53*, 981-992.
  47. Lerche, D.; Matsuzaki, S. Y.; Sørensen, P. B.; Carlsen, L.; Nielsen, O. J.; Ranking of Chemical Substances Based on the Japanese Pollutant Release and Transfer Register Using Partial Order Theory and Random Linear Extensions. *Chemosphere* **2004**, *55*, 1005-1025.
  48. Sørensen, P. B.; Mogensen, B. B.; Carlsen, L.; Thomsen, M. The Influence on Partial Order Ranking from Input Parameter Uncertainty: Definition of a Robustness Parameter. *Chemosphere* **2000**, *41*, 595-601.
  49. Brüggemann, R.; Carlsen, L. *Partial Order in Environmental Sciences and Chemistry*, Springer, **2006**.
  50. Rival, I. *Ordered Sets*: Proceedings of the NATO Advanced Study Institute held at Banff, Canada, Dordrecht Holland; Boston: D. Reidel Pub. Co.; Hingham, 1982.
  51. Rota, G. On the Foundation of Combinatorial Theory I, Theory of Möbius Functions. *Z. Wahr. Verw. Gebiete* **1964**, *2*, 340-368.